

PAPER • OPEN ACCESS

Generation of echocardiographic 2D images of the heart using cGAN

To cite this article: Vasily Zyuzin *et al* 2021 *J. Phys.: Conf. Ser.* **1727** 012013

View the [article online](#) for updates and enhancements.

 <p>The Electrochemical Society Advancing solid state & electrochemical science & technology 2021 Virtual Education</p> <p>Fundamentals of Electrochemistry: Basic Theory and Kinetic Methods Instructed by: Dr. James Noël Sun, Sept 19 & Mon, Sept 20 at 12h–15h ET</p> <p>Register early and save!</p>	
--	--

Generation of echocardiographic 2D images of the heart using cGAN

Vasily Zyuzin¹, Julia Komleva¹ and Sergey Porshnev^{1,2}

¹ Ural Federal University, IRIT-RTF, Russia

² N.N. Krasovskii Institute of Mathematics and Mechanics of the Ural Branch of the Russian Academy of Sciences, Russia

e-mail: v.v.zyuzin@urfu.ru

Abstract. One of the most significant tasks of echocardiography is the automatic delineation of the cardiac structures from 2D echocardiographic images. Over the past decades, the automation of this task has been the subject of intense research. One of the most effective approaches is based on the deep convolutional neural networks. Nonetheless, it is necessary to use echocardiogram frames of the cardiac muscle, which show the boundaries of the cardiac structures labeled/annotated by experts/cardiologists to train it. However, the number of databases containing the necessary information is relatively small. Therefore, generated echocardiogram frames are used to increase the amount of training samples. This process is based on the ultrasound images of the heart, annotated by experts. The article proposes an improved method for generating echocardiograms using a generative adversarial neural network (GAN) with a patch-based conditional discriminator. It has been demonstrated that it is possible to improve the quality of generated echocardiogram frames in both two and four chamber views (AP4C, AP2C) using the masks of cardiac segmentation with sub-pixel convolution layer (pixel shuffle). It is demonstrated that the proposed approach makes it possible to generate ultrasound images, the structure of which corresponds to the specified segmentation masks. It is expected that this method will improve the accuracy of solving the direct problem of automatic segmentation of the left ventricle.

1. Introduction

Echocardiography (EchoCG) is the primary means of measuring the cardiac morphology and function and to reach a diagnosis. EchoCG provides ultrasound images of the heart, which are used for identification of the edges of the cardiac structures.

Knowledge about the size of the cardiac structures allows to evaluate the quantitative characteristics used to diagnose the state of the heart muscle. The quantitative evaluation of the cardiac structures is necessary for a careful assessment and diagnosis. For instance, the estimation of the ejection fraction (EF) of the left ventricle (LV) requires an accurate delineation of the left ventricular endocardium in both end diastole (ED) and end systole (ES) parts of the cardiac cycle. In addition, using the geometric dimensions and shape of the left ventricle, it's possible to evaluate ejection fraction (EF), end-systolic volume (ESV) and end-diastolic volume (EDV) of a left ventricle. These indicators play a critical role in the diagnosis of heart diseases.

Currently, most modern software tools use manual or semi-automatic algorithms for delineation of the cardiac structures. The cardiologist annotates pivot points, and then a contour is drawn from them.



Content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](https://creativecommons.org/licenses/by/3.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

These methods turn out to be a laborious process and require relevant experience from cardiologists. At the same time, despite the existing guidelines of the American Society of Echocardiography [1], there are no strict rules for drawing the contours of cardiac structures on EchoCG images. This is due to the low quality of images and the presence of noise and the peculiarities of the ultrasound propagation in internal tissues. Therefore, the development of automatic algorithms for segmentation of heart areas is relevant.

To develop and confirm the possibility of using automatic algorithms, it is necessary to use datasets of annotated echocardiogram frames. The process of annotating ultrasound images requires the involvement of cardiologists and is very laborious. Therefore, the task of generating images from known masks of cardiac structures make sense.

This paper discusses the results of cGAN-based deep learning echo frames generation using the Cardiac Acquisitions for Multi-structure Ultrasound Segmentation (CAMUS) dataset, which contains 500 images of 50 patients in two and four chamber projections along the long axis of the LV.

2. Previous work for image generating

Convolution neural networks, in particular Generative adversarial networks (GANs), are traditionally used to solve a variety of computer vision problems, including:

- inpainting missing parts of the image [2–4];
- denoising [4, 5];
- image colorization [6, 7];
- style mixing [8, 9];
- semantic image manipulation [10, 11];
- increasing the image resolution [12].

The experience of using GAN to generate ultrasound images based on known masks of the internal regions of the heart is also known.

In [13], it is proposed to use two networks, one of which translates ultrasound images into sketch images, and the second, the generative one, learns the inverse mapping from the sketch image into corresponding echo frame. This generator is then used as a critic to assess and improve the segmentation mask generated by a given segmentation algorithm such as U-Net. This semi-supervised approach enforces a prior on the segmentation model based on the perceptual similarity of the generated frame with the original frame. This approach promotes utilization of the unlabeled samples, which, in turn, improves the segmentation accuracy.

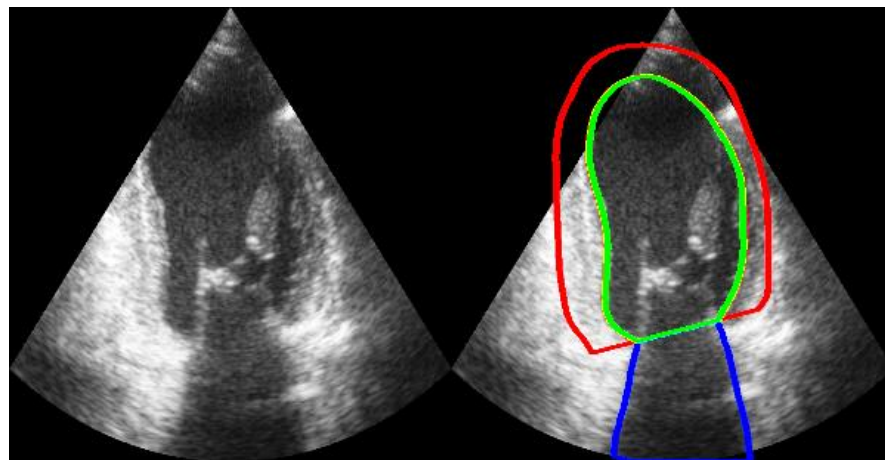
In [14], the architecture of a generative adversarial network with a conditional discriminator based on patches is proposed for generating echocardiograms of the CAMUS [15] dataset conditioned by segmentation mask. This approach allows generating 256×256 ultrasound images, however, the synthesized images do not contain speckle noise from the corresponding ultrasound frames, and a blurring and “checkerboard” artifacts appear.

Accordingly, the previously obtained results indicate the possibility of using the GAN for generating ultrasound images. However, it is necessary to modify the GAN in order to eliminate the disadvantages noted above.

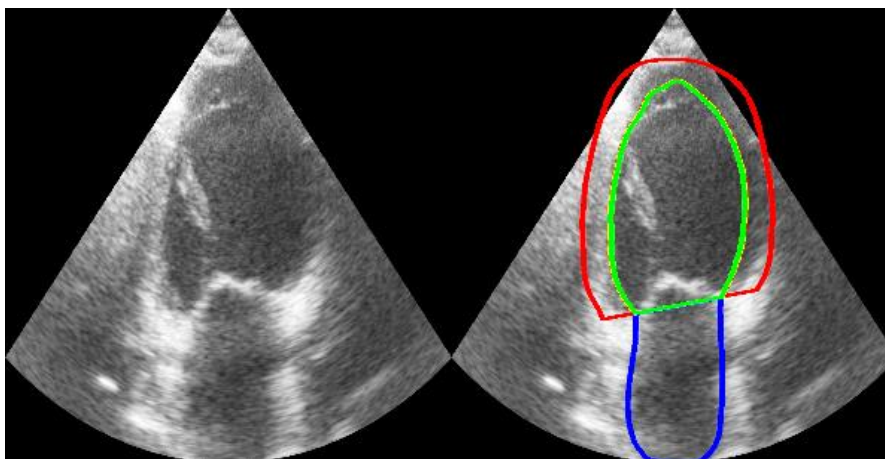
In our work, following [14], we used the CAMUS dataset, consisting of two-dimensional ultrasound images of the heart of 500 patients. For 450 patients, labeled frames are available in the open access: the area of the LV endocardium, the area of the LV epicardium and the area of the left atrium (LA). The test set consists of unlabeled echo frames of 50 patients. There are 2 images for each patient in a two-chamber and four-chamber view of the heart for the diastolic and systolic phases of the cardiac cycle. The total number of frames with expert annotation is 1800.

Analysis of image distributions according to expert estimates of image quality has shown that good quality has 916 frames (51%), medium quality – 682 frames (38%), poor quality – 202 frames (11%). Examples of typical frames are shown in Figure 1.

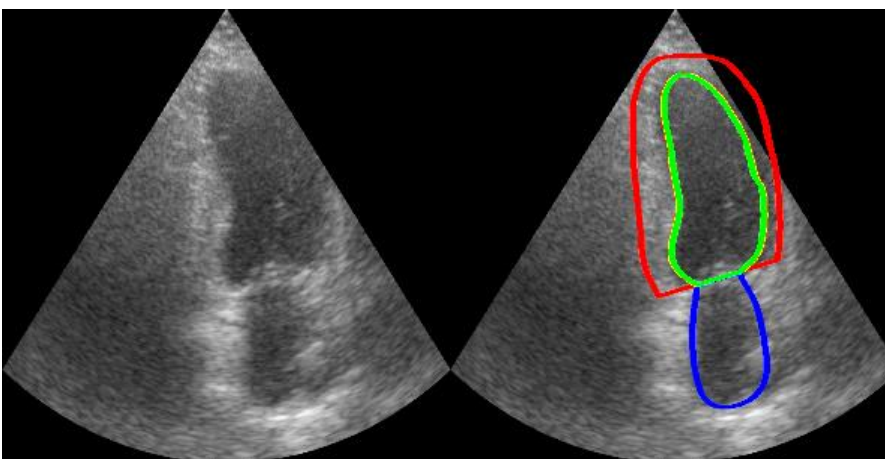
The distribution of ultrasound images of the dataset by their size is shown in Figure 2.



(a) Good quality



(b) Medium quality



(c) Poor quality

Figure 1. Typical echo frames with labeled cardiac structures

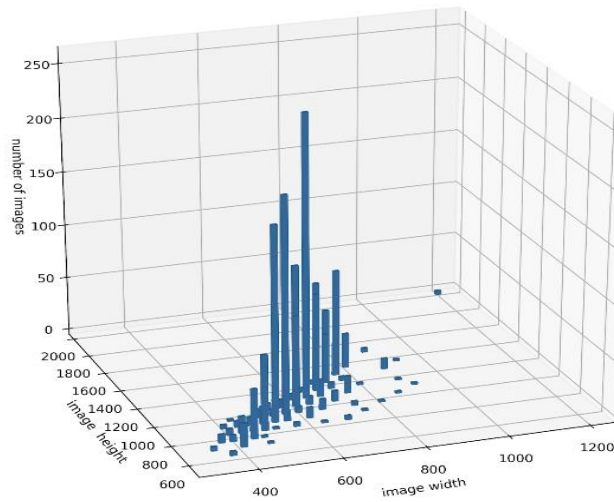


Figure 2. The distribution of ultrasound images of the dataset by their size

3. Modification of the cGAN architecture

The block diagram of the learning process used in [14] is shown in figure 3.

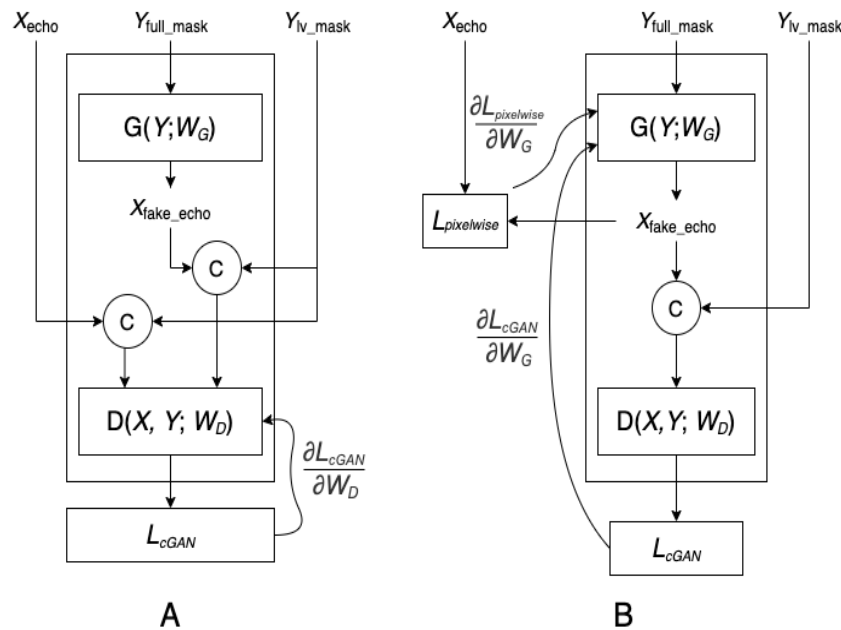


Figure 3. The learning diagram of the cGAN's discriminator (A) and generator (B):

X_{echo} – echo frames, $X_{\text{fake echo}}$ – generated echo frames, $Y_{\text{full mask}}$ – all segmentation masks (LV endocardium, LV epicardium, LA), $Y_{\text{lv mask}}$ – LV segmentation mask, C – concatenation of images, WG – generator weights, WD – generator weights, $L_{\text{pixelwise}}$ – loss function for comparison X_{echo} and $X_{\text{fake echo}}$, L_{cGAN} – loss function of cGAN

Figure 3 shows that the learning process consists of two sequential stages:

- training D;
- training G.

G generates ultrasound images based on expert masks, which are then compared with the corresponding real ultrasound images. According to the criterion $L_{\text{pixelwise}}$ (mean absolute error) the G weights are being changed. The synthesized frames, real frames and the corresponding LV masks are then passed to D, which classifies whether the image patches are real or generated. Criterion in the conventional cGAN L_{cGAN} is used mean square error (MSE) as adversarial loss for changing weights of G and D.

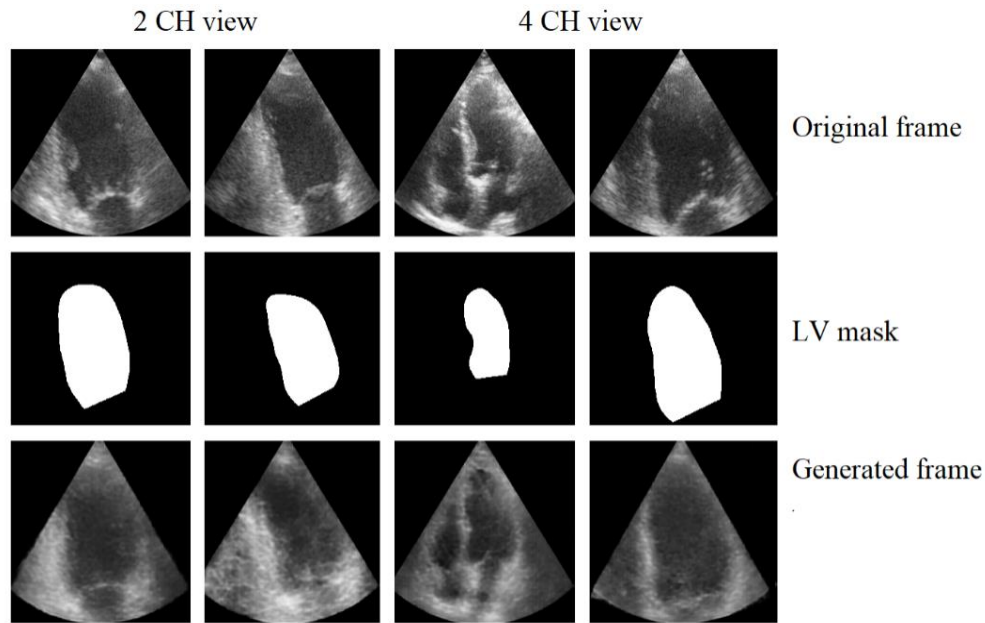


Figure 4. Typical outputs of the discussed cGAN

The figure shows that the synthesized image has a checkerboard artifact [16], and the synthesized ultrasound images do not have speckle noise, which is an integral part of real ultrasound images. The reason for these problems is associated with the use of transposed convolutions to increase the size of the image, implemented in G in [14]. Let us consider in more detail the implementation of the structure of G used in [14]. The G is based on the standard Unet architecture [17], and consists of 7 convolution layers and 7 deconvolution layers without skip-connections. The deconvolutions were implemented as transposed convolutions. Outputs of all convolutional and deconvolutional layers were batchnormalized, except for the last layers of D and G, and followed by the Leaky ReLU activations. In this case a 4×4 convolution kernel was used in each G block. In connection with the above, a hypothesis was put forward that to get rid of the checkerboard artifacts and obtain, first, less blurry images with a more pronounced texture of the echocardiogram, secondly, increase the size of the generated images to 512×512 .

To confirm the stated hypothesis, skip-connections between encoder and decoder blocks were implemented in the G architecture. Transposed convolutions layers were replaced with a convolution layer with kernel size 3×3 , stride 1, padding 1, which increases in the number of feature maps by 4 times, and sub-pixel convolution layer (pixel shuffle), which raises the resolution of low-level features by 2 times.

4. Results

Let us consider the results of a comparative analysis of synthesized ultrasound images using G in [14] and our modified G. Here G and D were trained using the Adam optimizer with learning rates 0.00013 and 0.00015, respectively. The batch size was 32. The size of the discriminator patch was 16×16 pixels. The weight of the adversarial term was set to $\lambda = 0.01$. The training took place over 500 epochs. Typical LV ultrasound images generated by modified G are shown in Figure 7.

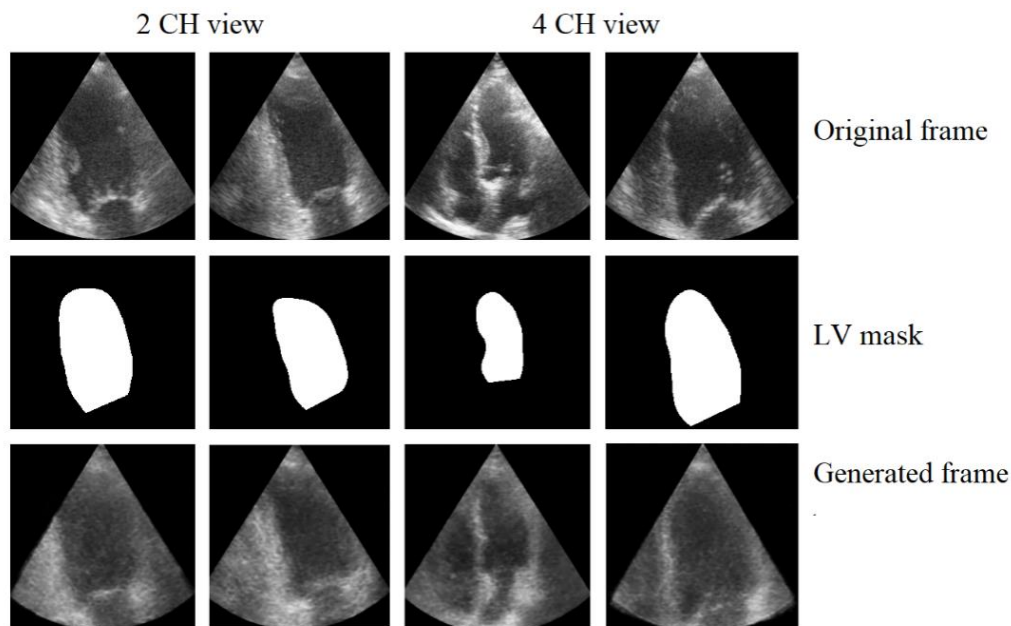


Figure 5. Typical outputs of the modified cGAN

Figure 7 shows that the proposed modification of the G cGAN, in fact, made it possible to reduce the checkerboard effect.

Results are reported in Table 1. For a quantitative comparison of the generated ultrasound images the peak signal-to-noise ratio (PSNR) and the structural similarity index (SSIM) are used and presented in Table 1.

Table 1. PSNR and SSIM estimates.

	PSNR smoothed		SSIM smoothed	
	2ch	4ch	2ch	4ch
cGAN [14] (256×256)	8.404	8.387	0.5569	0.5498
Our cGAN (256×256)	8.411	8.407	0.5694	0.5680
Our cGAN (512×512)	8.396	8.389	0.6033	0.6002

Table 1 shows that the GAN has learned to map from segmentation masks to the corresponding cardiac structures. As a result, the model can generate a 512×512 long-axis two-chamber and four-chamber view.

5. Conclusions and future work

A cGAN generator modification for the generation of ultrasound images based on expert LV masks in two and four-chamber heart views is proposed.

The results confirm that the modified cGAN generator makes it possible to reduce the "checkerboard" and blurring effects on generated ultrasound images, as well as to ensure the similarity of their structure to real ultrasound images.

In further studies, the authors plan to develop an approach that provides the generation of echo frames of various quality required to test the applicability of automatic algorithms for cardiac segmentation.

The work was supported by Act 211 Government of the Russian Federation, contract № 02.A03.21.0006.

References

- [1] Lang R, Badano L, Mor-Avi V et al 2015 *EACVI* **16(3)** 233
- [2] Yu J, Lin Z, Yang J, Shen X, Lu X and Huang T 2018 *CVPR* 5505
- [3] Liu G, Reda F, Shih K, Wang T, Tao A and Catanzaro B 2018 *ECCV* 85
- [4] Xie J, Xu L and Chen E 2012 *NIPS* 341
- [5] Zhang K, Zuo W, Chen Y, Meng D and Zhang L 2017 *IEEE Trans. Image Process.* **26(7)** 3142
- [6] Zhang R, Zhu J, Isola P, Geng X, Lin A, Yu T and Efros A 2019 arXiv (*preprint*:1705.02999)
- [7] Nazeri K, Ng E and Ebrahimi M 2018 *AMDO* 85
- [8] Semmo A, Isenberg T and Döllner J 2017 *NPAR* 1
- [9] Karras T, Laine S and Aila T 2019 *CVPR* 4401
- [10] Hong S, Yan X, Huang T and Lee H 2018 *NIPS* 2708
- [11] Lee C, Liu Z, Wu L and Luo Maskgan P 2020 *CVPR* 5549
- [12] Johnson J, Alahi A and Fei-Fei L 2016 *ECCV* 694
- [13] Jafari M, Girgis H, Abdi A, Liao Z, Pesteie M, Rohling R and Abolmaesumi P 2019 *ISBI* 649
- [14] Abdi A, Tsang T and Abolmaesumi P 2019 arXiv (*preprint*:1911.02121)
- [15] Leclerc S, Smistad E, Pedrosa J, Østvik A, Cervenansky F, Espinosa F and Lartizien C 2019 *IEEE Trans Med Imaging* **38(9)** 2198
- [16] Aitken A, Ledig C, Theis L, Caballero J, Wang Z and Shi W 2017 arXiv (*preprint*:1707.02937)
- [17] Ronneberger O, Fischer P and Brox T 2015 *MICCAI* 234